情報幾何とその応用(脳化学2,数学者のための分子生物学入門-新しい数学を造ろう-)

AUTHOR(S):

甘利, 俊一; 藤原, 祐介

CITATION:

甘利, 俊一 ...[et al]. 情報幾何とその応用(脳化学2,数学者のための分子生物学入門-新しい数学を造ろう-). 物性研究 2006, 87(3): 457-466

ISSUE DATE:

2006-12-20

URL:

http://hdl.handle.net/2433/110689

情報幾何とその応用

甘利俊一 (理化学研究所 脳科学総合研究センター)

レクチャーノート作成　藤原祐介（奈良先端科学技術大学院大学）

概要

はじめに情報幾何の基礎を紹介する．続いて，情報幾何の応用例として，線形関係の推定，神経細胞の集団符号化とその相互作用，多層パーセプトロンと特異モデルを順に紹介していく．

1　導入 [1,7–9]

情報幾何の基礎について簡単に述べる．

1.1　確率分布の作る多様体と２つの不変性

確率分布の全体を考えよう．例えば確率変数 x の正規分布

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

の全体を考えよう．パラメータは μ と σ の２つなので，正規分布の全体は二次元の多様体になっている．この多様体の局所構造はどうなっているのだろうか．局所構造を調べるためにリーマン計量などを自然に入れる方法がないだろうか．

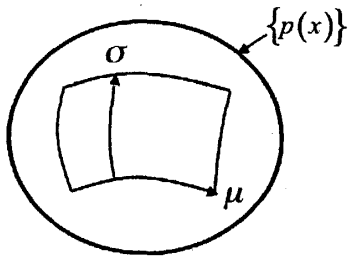

図 1

もっと単純な離散確率分布を考えてみよう．確率変数 x は 1,2,3 の値をとる離散確率変数とする．確率ベクトルは

$$\boldsymbol{p} = (p_1, p_2, p_3)$$

と表され，確率は足して 1 になるから

$$p_1 + p_2 + p_3 = 1$$

である．

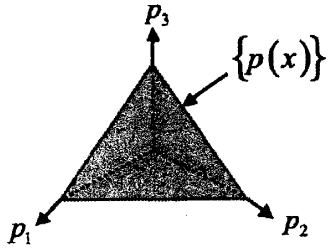

図 2

このため確率ベクトル $\boldsymbol{p}$ の全体がつくる多様体は２次元になる．図にすると３角形状の空間になっている．この上に計量が定義できたとすれば，ある確率分布ともうひとつの確率分布がどれだけ近いを測るものさしとなる．そのものさしは確率に則して決めなければならない．３角形だからといってユークリッド空間で良いかというとそうではない．例えば，$\xi_i = \sqrt{p_i}$ とパラメータをとればこの分布のなす空間は球面になる．この場合はユーク

リッド幾何ではなく球面の幾何学が必要となる．この他にもパラメトリゼーションの仕方はいくつもある．$S = \{p(x,\theta)\}$ という確率分布の空間を考える場合に，パラメータ θ の取り方に依存しない幾何構造（計量）を決めたい．

また $y = y(x)$ という非線形関数があって，確率変数 x を y に変換したとしよう．y の確率分布になったとしても持っている情報は x と変わらない．だから x で定義しようが y で定義しようが同じ幾何構造を与えるものを考えなければならない．したがって次の２つの不変性が重要である．

1. パラメトリゼーションの不変性
2. 確率変数の表現の不変性

この２つの条件だけで確率分布の局所的な幾何構造が一意的に決まると予想されている．

1.2 リーマン構造と双対接続

リーマン構造はどうなるのだろうか．リーマン計量は

$$ds^2 = \sum g_{ij} d\theta_i d\theta_j$$

で与えられる．このときテンソル g_{ij} が

$$g_{ij}(\boldsymbol{\theta}) = E\left[\frac{\partial}{\partial\theta_i}\log p \ \frac{\partial}{\partial\theta_j}\log p\right]$$

で与えられるとき，リーマン構造が一意に決まる．$g_{ij}(\boldsymbol{\theta})$ は Fisher 情報行列と呼ばれている．

リーマン構造が決まると，アファイン接続は，接続が対称であるという条件のもとで，一意的にリーマン接続となる．リーマン接続とは平行移動しても内積が変わらない接続である．しかし，これだけだとほとんど何も発展性がない．それならば他に構造を考えてみよう．リーマン接続は計量を保存した接続であったが，計量と接続を別々に定義することもできるのではないか．しかし，これではリーマン構造を壊してしまう恐れがある．リーマン構造を崩さないで議論を豊かにするには接続を２つ定義するのがよい．∇ とそれに双対な ∇^* $((\nabla^*)^* = \nabla)$ である．もしくは平行移動で定義した Π と Π^* である．２つのベクトル場 X と Y があったときに，その内積が

$$\langle X, Y\rangle = \langle \Pi X, \Pi^* Y\rangle$$

で保存される構造を考える．このとき重要なのは３つの値 (g, ∇, ∇^*) であり，この３つが双対構造を決める．$\nabla = \nabla^*$ ならば自己双対でリーマン接続となる．ここでは捩率は考えない．もし ∇ と ∇^* が平坦なとき，(g, ∇, ∇^*) によって決められる空間を双対平坦空間と呼ぶ．双対平坦空間では，座標系 $\boldsymbol{\theta}$ と双対な座標系 $\boldsymbol{\eta}$ は，ある滑らかな凸関数 $\phi(\boldsymbol{\theta})$ が与えられたときにルジャンドル変換:

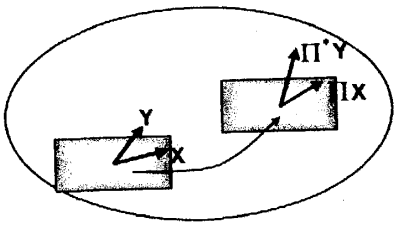

図３

$$\boldsymbol{\eta} = \frac{\partial}{\partial\boldsymbol{\theta}}\phi(\boldsymbol{\theta})$$

で結ばれる．見方を変えると，凸関数のルジャンドル変換における幾何学構造は双対平坦であるといえる．

1.3　ダイバージェンスと統計的推定

双対平坦空間ではどのような良いことがあるのだろうか．実は双対平坦性から2点間 p, q のダイバージェンス:

$$D[p:q]$$

という量が自動的に決まる．さらに，3点 p, q, r に対して，p と q を結ぶ線を測地線（m-測地線），q と r を結ぶ線を双対測地線（e-測地線）として，この二つが直交したときにピタゴラスの定理が働き

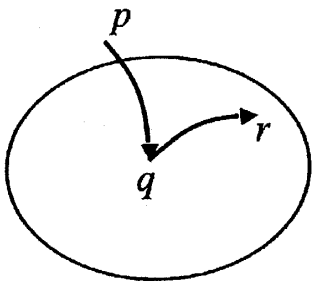

図4

$$D[p:r] = D[p:q] + D[q:r]$$

となる．このとき射影定理が成立する．ある点からある部分多様体上の一番近い点を探したいとき，射影すれば良いのである．確率分布 $p(x)$ の作る多様体では，一つの測地線は確率分布の mixture

$$r(x,t) = tp(x) + (1-t)q(t)$$

で表され，もう一つの測地線は対数をとった曲線

$$\log r(x,t) = t\log p(x) + (1-t)\log q(t)$$

で表される．前者を m-測地線，後者を e-測地線と呼ぶ．

統計学ではデータが測定されたときにそれが満たす確率分布を推定する．これは有限次元のパラメータで指定された分布ならばそのパラメータを推定することと等しい．データが多くなればなるほど推定精度は良くなる．この場合，推定値の候補が，データから確率分布の作る多様体に射影した脚の近傍だけに限られるので，接空間の議論になる．これは1次漸近論と呼ばれ，どのような確率分布でも同じように扱うことができる．さらにデータがもう少し広がりをもち2次近似が必要になる場合には曲率を考えればよい．

2　最小2乗法は嘘っぱち? [5,6]

2.1　最小2乗法

n 個の観測値のペア $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ があり，x と y には原点を通る線形な関係

$$y = \theta x$$

がある．このとき θ を推定したい．ただし，x_i は真のデータ ξ_i に誤差 ϵ_i が上乗せされて観測されたもの

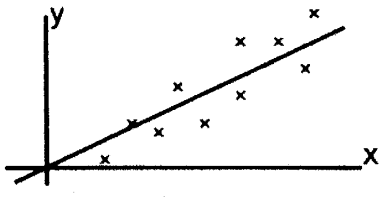

図5

$$x_i = \xi_i + \epsilon_i$$

とする．y_i も同様に誤差 ϵ_i' を含む:

$$y_i = \theta\xi_i + \epsilon_i'.$$

簡単のため，誤差 ϵ_i, ϵ_i' はすべて互いに独立で平均0分散 σ^2 の正規分布 $N(0, \sigma^2)$ に従うとする．

最小２乗法は，

$$L(\theta) = \sum (y_i - \theta x_i)^2$$

を最小にする θ を求める．解は，

$$\hat{\theta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

である．この答えはそう悪くはないが，一番良い方法ではない．

2.2 その他の方法

この他にも答えは考えられる．例えば，

$$\frac{1}{n}\sum \frac{y_i}{x_i} \quad や \quad \frac{\sum y_i}{\sum x_i}$$

などである．

良さそうな方法は，ある仮想の直線を考えて，それに観測値を正射影したときの誤差の２乗和を最小にする $\hat{\theta}$ を求める方法である．なぜなら，x の誤差も y の誤差と同じように扱っているからである．けれども，この方法と一番単純な答え $\sum y_i / \sum x_i$ とどちらがよいだろうか．実はこれは場合による．

なぜこのような変なことがおこるのだろうか．

2.3 セミパラメトリックモデル

今，θ は未知である．そして，$\xi_1, \cdots, \xi_n$ も未知なのである．つまり，未知パラメータが $n+1$ 個ある．だから，データをとってもとっても未知パラメータが増えていくのである．この問題は 1950 年代から知られており Neyman-Scott 問題と呼ばれている．今の問題の場合，未知パラメータが (θ, ξ) であるときの観測値 (x, y) の確率分布は

$$p(x, y|\theta, \xi) = c \exp\left\{ -\frac{1}{2}(x - \xi)^2 - \frac{1}{2}(y - \theta\xi)^2 \right\}$$

と書ける．未知パラメータ ξ が未知の分布 $Z(\xi)$ から独立に生じたものであるとすると，(x, y) の同時確率分布は

$$p(x, y|\theta) = \int p(x, y|\theta, \xi) Z(\xi) d\xi$$

と書ける．$p(x, y|\theta)$ は未知の関数 $Z()$ と未知のパラメータ θ を含んでいる．この統計モデルをセミパラメトリックモデルという．

この場合の推定論はどうしたらいいのだろうか．実は情報幾何で扱うことができるのである．関数空間において，パラメトリックモデルの上にファイバーバンドル的なものを考えて，求めたい推定関数がパラメトリックな関数と直交するように推定すればよい．

ニューロンのパルス間隔の構造の推定についても，同様の推定論を用いて良い推定ができる．一つのパルスが出た後に次のパルスが出るまでの間隔が独立ならば，パルス間隔は指数分布になる．規則的に出るならば，

一定間隔である．実際のニューロンはそのどちらでもない．実験データは発火率が測定ごとに揺らいでいる．発火率を知りたいわけではなく，パスル間隔の揺らぎを規定するパラメータを知りたい．この問題はガンマ分布のセミパラメトリックモデルとして取り扱うことができる．この問題についても情報幾何の観点からよい推定方法が得られている．

3　神経細胞の集団符号化 [2,10]

3.1　ニューロンの相互作用

ニューロンが n 個あって，確率的に発火したりしなかったりする．それぞれのニューロンの状態を n 個の変数 $x_1, ..., x_n$ で表す．発火していれば $x_i = 1$，発火していなければ $x_i = 0$ とする．発火する確率は条件によって変化する．このとき n 個のニューロンの状態の同時確率分布

$$p(\mathbf{x}) = p(x_1, x_2, \cdots, x_n)$$

を議論したい．

まず，この確率分布からわかる一番簡単な値は x_i の期待値

$$\eta_i = E[x_i],$$

つまり発火率である．また，ニューロン間には相互作用があるはずで，その度合いを知りたい．その値として例えば，x_i と x_j の共分散

$$v_{ij} = Cov[x_i, x_j]$$

が考えられる．

問い　相互作用を調べるにはどのような値がよいのだろうか，相関係数が良いのだろうか．

まず，二つのニューロンだけを考えよう．二つのニューロンの状態を x_1, x_2 とすると，その同時確率分布 $p(x_1, x_2)$ は 3 次元空間 $S = \{p(x_1, x_2)\}$ をなす．x_1, x_2 が独立なときの確率分布がなす部分空間は 2 次元曲面 $M = \{p(x_1)p(x_2)\}$ である．二つのニューロンに相互作用がある場合，p はこの曲面からはずれている．そのとき，相互作用を表すための座標は M と直交している方向にとれば良い．直交しているようにとればそれぞれのニューロンの状態とは関係なく相互作用のみを測ることができるからである．直交している座標は

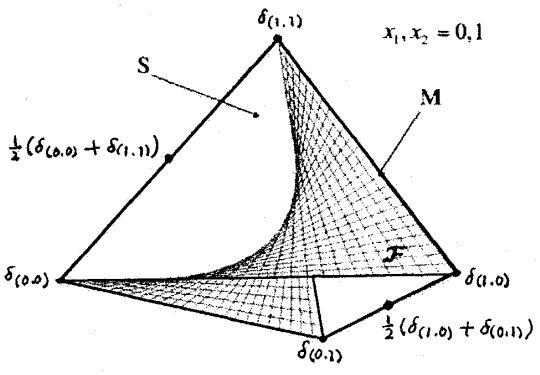

図 6

$$\theta = \log \frac{p(1,1)p(0,0)}{p(1,0)p(0,1)}$$

で与えられる．独立な場合，$p()$ が積の形に分解されるので $\theta = 0$ となる．これを相互作用を測る座標とすればよい．

3.2 3次の相互作用

次に3つのニューロンの場合を考えよう．2つのニューロン間には相互作用が無い場合でも，全体として独立かというとそうではない．3次相関を考える必要がある．そのためには，2次相関まではあって3次相関がないという部分空間と，1次，2次相関は無くて3次相関のみがあるとう部分空間に分けて考えると見通しが良い．情報幾何の観点から見ると，この二つの部分空間は直交していることがわかっている．さらに高次の相関についても同様に直交分解を考えることができる．この性質は，昔から統計学では現象として知られていたが，情報幾何によって直交分解になっていることがわかったのである．

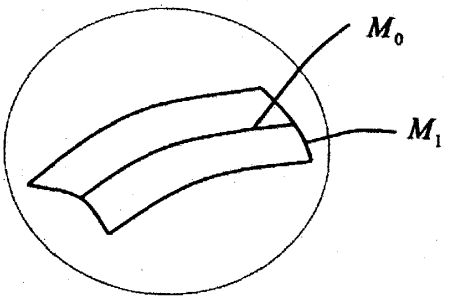

図7

3.3 高次の相互作用と同期発火

最後にn個ニューロンがある場合を考えよう．そうすると全部で 2^n-1 の状態を持つことになる．高次相関を考えたいが，nが大きい場合，その値を決めるのは不可能に近い．そこでなにかニューロン間に構造をいれなければならない．その中の一つとして，それぞれのニューロンがある共通の入力をうけて発火するモデルを考える．そのとき高次相関がどういう現象として表れるかを調べてみよう．共通の入力によって発火するのだから，ニューロン間の発火率には相関がある．このときn個のニューロンのうち何個が同時に発火しているかの割合

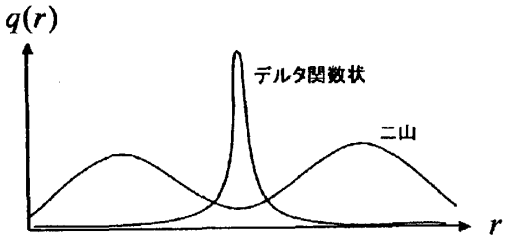

図8

$$r=\frac{1}{n}\sum x_i$$

の確率分布 $q(r)$ を求めたい．すべてのニューロンの発火率が独立で同一な分布に従うと仮定すれば，発火率は大数の法則よりデルタ関数状になる．2次相関はあって、3次以上の相関が無い場合も、相関のある場合の大数の法則によりデルタ関数状になる。ニューロンの集団が同期発火する場合，つまり高次相関がある場合には二山になる。この二山の分布は，あるときには多くのニューロンが同時に発火するが，あるときにはほとんど発火しないということを表す．

3.4 同期発火とバインディング問題

ニューロンの集団が同期して発火すると，そのときに運ばれる情報は大きいから，脳のなかでは同期して発火するという現象が次々に起こっているのではないかといわれている．実際，ニューロン集団の同期現象がバインディング問題に関連があるのではないかと考えられている．今，目の前に赤い○と黒い□があったとしよう．脳の中には形を認識するニューロン集団があって，○に反応するニューロンが発火し，同時に□に反応するニューロンも発火する．また，色を認識するニューロン集団があって，赤に反応するニューロンと黒に反応するニューロンも同時に発火する．これらのニューロンが同時に発火したら，○が赤いのか□が赤いのか区別

がつかないのではないだろうか．しかし，私たちは簡単に見分けることができる．それでは脳ではどのような情報の交通整理が行われているのだろうか．これがバインディング問題である．一つの説は，○のニューロンの発火と赤のニューロンの発火が同期しているのではないかというものである．

4　多層パーセプトロンの学習と特異モデル [3,4,11]

人工神経回路網の一種，多層パーセプトロンのつくる関数全体を関数空間上の多様体だと考え，その学習が多様体の幾何学構造を考慮した自然勾配法を用いることにより大きく改善される．

4.1　多層パーセプトロン

まず，多層パーセプトロンのモデルについて説明する．n 個の入力信号 $\boldsymbol{x}=(x_1,\cdots,x_n)$ が与えられて，それを m 個の中間層のニューロンが受け取る．j 番目の中間層のニューロンは入力信号に重み $\boldsymbol{w}_j$ をかけた値を非線形関数 ψ で変換した

$$z_j=\psi(\boldsymbol{w}_j\cdot\boldsymbol{x})$$

を出力する．最後のニューロンは中間層の出力に重み $\boldsymbol{v}$ をかけて

$$y=\sum_i v_i\psi(\boldsymbol{w}_i\cdot\boldsymbol{x})$$

を出力する．ここで，すべての重みパラメータをまとめて

$$\boldsymbol{\theta}=(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_m;\boldsymbol{v})$$

とする．そして出力にノイズ ϵ が加わるとすると，出力関数は

$$y=\sum_i v_i\psi(\boldsymbol{w}_i\cdot\boldsymbol{x})+\epsilon=f(\boldsymbol{x},\boldsymbol{\theta})+\epsilon$$

と書ける．ノイズ ϵ は平均 0 分散 1 の正規分布に従う確率変数とする．このとき $\boldsymbol{x}$ が与えられたときの y の確率分布は，

$$p(y|\boldsymbol{x};\theta)=c\exp\left\{-\frac{1}{2}(y-f(\boldsymbol{x},\boldsymbol{\theta}))^2\right\}$$

と表される．

問題はパラメータ $\boldsymbol{\theta}$ を決めることである．パラメータの学習は，ある $\boldsymbol{x}$ が与えられたときに，それとあわせて正解 y が与えられることによって行われる．つまり，ある $\boldsymbol{x}$ に対しては出力は y になるべきであるという例題集 $\{(y_1,\boldsymbol{x}_1),\cdots,(y_n,\boldsymbol{x}_n)\}$ が与えられて，逐次的に学習が行われるのである．ただし y にはノイズがのるので必ずしも与えられる教師信号は正しくない．結局，これはデータ $(y,\boldsymbol{x})$ が与えられたときのパラメータ $\boldsymbol{\theta}$ を求める統計的推定の問題である．

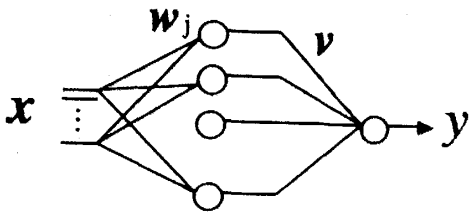

図 9

4.2　ニューロ多様体

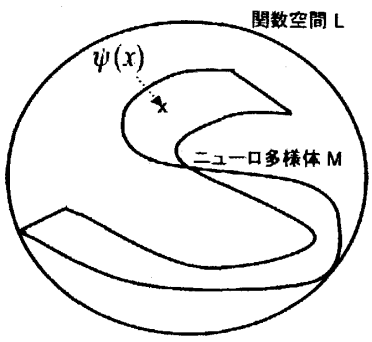

図 10

一般の非線形関数 $\psi(\boldsymbol{x})$ が作る関数空間 L を考えよう．パラメータ $\boldsymbol{\theta}$ の次元を N とすると，パーセプトロン全体がつくる関数の空間は，関数空間 L の中の N 次元部分空間となる．この部分空間 M をニューロ多様体とよぶ．パーセプトロンが処理できるものはニューロ多様体 M の中にあるものだけで，その外にあるものは処理できない．本当のデータが M の外にある関数から出てきた場合はパーセプトロンは M 上で最もその関数に近い点を探す．学習というのは，例題からその点に至るようにパラメータを変化させることである．ただ，M が曲がっている場合，遠回りをしなければならない．これが勾配法で学習が非常に遅くなるひとつの原因である．では M がまっすぐならばいいのだろうか．しかし，そうではない．まっすぐだと様々な関数をカバーできなくなってしまう．曲がっている方が情報処理を行える容量が大きいのである．そのため，ある関数族の情報処理の容量とその学習の効率を議論するときには幾何学的な構造を考える必要がある．今，パーセプトロンがつくる関数が確率分布族だとすると M は双対接続の空間になっている．けれども，この場合，ノイズが正規分布に従うとしているので，双対性が消えリーマン空間になる．

4.3 パーセプトロンの学習

学習はどうするかというと，今，パラメータの値が $\boldsymbol{\theta}_t$ であるとき，入力 $\boldsymbol{x}_t$ によって得られる出力 $f(\boldsymbol{x}_t, \boldsymbol{\theta})$ と，正解 y_t との誤差が小さくなるように新たな $\boldsymbol{\theta}_{t+1}$ を推定する．その 2 乗誤差は，

$$l(y_t, \boldsymbol{x}_t; \boldsymbol{\theta}_t) = \frac{1}{2}\left|y_t - f(\boldsymbol{x}_t, \boldsymbol{\theta}_t)\right|^2 = -\log p(y_t, \boldsymbol{x}_t; \boldsymbol{\theta}_t)$$

である．通常の勾配法による学習は，誤差関数 l の勾配をとって

$$\Delta\boldsymbol{\theta}_t = -\eta_t \nabla l$$
$$\nabla l = \left(\frac{\partial l}{\partial \theta_1}, \cdots, \frac{\partial l}{\partial \theta_N}\right)$$

である．この学習方法はバックプロパゲーションと呼ばれている．

4.4 特異構造

ニューロ多様体は θ だけをとって見たならばユークリッド空間に見えるが，関数として見たとき，いったいどうのような構造をしているのだろうか．構造を調べるためには計量を計算すれば良い．パーセプトロンの作る空間の計量は Fisher 情報行列

$$g_{ij}(\boldsymbol{\theta}) = E\left[\frac{\partial \log p(\boldsymbol{x}, y; \boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \log p(\boldsymbol{x}, y; \boldsymbol{\theta})}{\partial \theta_j}\right]$$

で与えられる．

では，本当に M は多様体なのだろうか，ところどころに特異点があるのではないだろうか．

その前に，なぜ関数空間に特異点が生じるか簡単なモデル

$$y = \xi\psi(\boldsymbol{w}\cdot\boldsymbol{x}) + \epsilon$$

を例にとって考えてみよう．これは多層パーセプトロンから一つの中間層のニューロンを取り出してきたものである．$\boldsymbol{w}$ と ξ を軸にとったのが図 11 の A である．$\xi = 0$ のとき $\boldsymbol{w}$ はどのような値でも同じ関数になる．つまり，この空間は一見，A のように見えるが，実際は B のような $\xi = 0$ で一点に縮んだ形をしている．この点が特異点である．

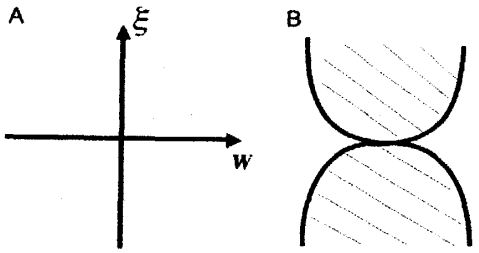

図 11

実際に多層パーセプトロンでも特異点がいくつも生じている．なぜかというと，特異点が生じる現象は対称性に由来しているからである．多層パーセプトロンはニューロンがたくさんあり，ニューロン i とニューロン j の重みが $\boldsymbol{w}_i = \boldsymbol{w}_j$ となって 2 つのニューロンが同じ動作をするようになってしまう場合が多々ある．こうなると 2 つニューロン間の責任分担が任意になるので特異点が生じてしまうのである．

4.5　自然勾配法

通常の最急降下法は，ユークリッド空間でかつ直交座標系をとっているときには最急降下方向に勾配をとる．しかし，リーマン空間では最急降下方向にはならない．とくに特異点がある空間では，特異点にトラップされてしまって学習がなかなか進まなくなってしまう．

通常の最急降下法は

$$\Delta\boldsymbol{\theta}_t = -\eta_t \nabla l$$

であった．左辺は反変ベクトル，右辺は共変ベクトルであり，左辺と右辺の性質が異なる．右辺を左辺に合わせるためには計量行列を使えば良い．リーマン空間での最急降下方向は

$$\tilde{\nabla} l = G^{-1}(\boldsymbol{\theta})\nabla l$$

で与えられる．これを自然勾配と呼ぶ．$G(\boldsymbol{\theta})$ は Fisher 情報行列である．よって学習法則は

$$\Delta\boldsymbol{\theta}_t = -\eta_t \tilde{\nabla} l$$

とするのが良い。これを自然勾配法と呼ぶ．

しかし，自然勾配を求めるには，Fisher 情報行列を計算してその逆行列をとらなければならない．この計算は結構大変である．そこで，G_t^{-1} をデータから逐次的に推定していく方法

$$\hat{G}_{t+1}^{-1} = (1+\epsilon)G_t^{-1} - \epsilon\tilde{\nabla} l_t \left(\tilde{\nabla} l_t\right)^T$$

を用いて計算すれば，それほど計算量は多くならずにすむ．この適応的自然勾配法（点線）と，通常の勾配法（実線）の学習を比べたのが図 12 である．通常の勾配法は学習の途中で特異点に引き込まれてプラトーに陥っている．これは特異点で ∇l が 0 になっているからである．一方，自然勾配法は，特異点で G^{-1} が無限大になり，自然勾配 $G^{-1}\nabla l$ は適切な値をとるので，プラトーに陥ることなく難なく収束するのである．

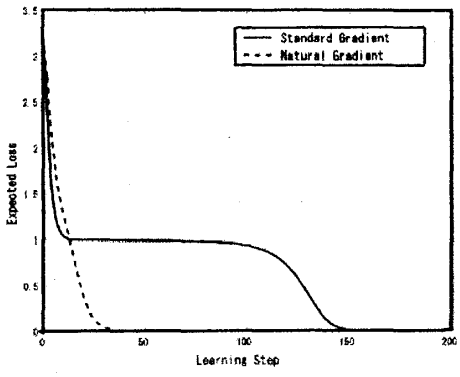

図 12

参考文献

[1] S. Amari and H. Nagaoka, Methods of Information Geometry, American Mathematical Society and Oxford University Press (2000)

[2] H. Nakahara and S. Amari, Information-Geometric Measure for Neural Spikes, Neural Comput. Vol.14, pp.2269-2316, (2002)

[3] S. Amari, H. Park, K. Fukumizu, Adaptive Method of Realizing Natural Gradient Learning for Multilayer Perceptrons Neural Comput. Vol.12, pp.1399-1409, (2000)

[4] S. Amari, H. Park, T. Ozeki, Singularities Affect Dynamics of Learning in Neuromanifolds Neural Comput. Vol.18, pp.1007-1065, (2006)

[5] 甘利俊一，川鍋元明，線形関係の推定ー最小２乗法は最良であるのか？，応用数理, Vol.6, pp.96-109, (1996)

[6] 三浦佳二，変動する環境のもとでのスパイク不規則性パラメタの不偏推定，京都大学博士論文，(2006)

[7] 甘利俊一，情報幾何とその応用 1 情報幾何とは何かー入門編，システム／制御／情報, Vol.48, No.6, pp.227-235, (2004)

[8] 甘利俊一，情報幾何とその応用 2 凸解析と双対平坦空間，システム／制御／情報, Vol.48, No.8, pp.340-347, (2004)

[9] 甘利俊一，情報幾何とその応用 3 統計的推論の情報幾何，システム／制御／情報, Vol.48, No.10, pp.428-436, (2004)

[10] 甘利俊一，情報幾何とその応用 7 神経集団符号化と高次相互作用，システム／制御／情報, Vol.49, No.6, pp.238-245, (2005)

[11] 甘利俊一，情報幾何とその応用 8 神経多様体における学習と特異モデル，システム／制御／情報, Vol.49, No.8, pp.337-343, (2005)